Utilizing On-Board GPS in City Buses to Determine Traffic Conditions

Ikhwan Hafiz Muhamad and Ng Siew Ling
Faculty of Electrical & Electronics Engineering, Universiti Malaysia Pahang, 26600 Pekan, Pahang, Malaysia
ikhwanh@ump.edu.my

***Abstract*—The traffic congestion has become a major concern to the society. It causes difficulties in journey planning while avoiding the traffic congestion. Increasing number of vehicles lead to traffic congestion, especially during peak hours. There are many Mobile Applications (App) which can update traffic conditions in certain routes, but these applications such as Waze requires the road users to manually update the traffic conditions. Besides that, the road users also require to be online to get the real-time traffic conditions. On the other hand, the traffic data of this App also will not be accurate if fewer people are using this app on that route. Therefore, this project aims to provide automatic updates on traffic conditions to each road user without the need of installing additional App and updates from the user. The traffic conditions are predicted using the onboard GPS data in the city buses. The traffic monitoring algorithm is developed using the Fuzzy Logic Algorithm. The result is displayed in a Graphical User Interface (GUI) and a push notification to the user's smartphone. The accuracy of this system is 90.34% where the inaccurate data occurred mostly in the data at Pekan, Pahang area due to the unexpected road conditions such as the deflection of the road, uneven road, holes and wild animals crossing which cause the bus driver to slow down the speed.**

***Index Terms*— Automatic Traffic Monitoring; Bus GPS; Fuzzy Logic Algorithm; Mobile Application.**

I. INTRODUCTION

Smartphone, satellites and sensor technologies had become more advanced and used to improve the efficiency of intelligent transportation systems (ITS) of existing transport system [1]. Traffic prediction is the most important criteria to be focused on since the traffic congestion problems had become a major concern from days to days as the traffic and cities had developed together. Traffic congestion is characterized by slower speeds, longer trip times, and increased queueing of vehicles on the road. The traffic congestion will reduce productivity and increase fuel consumption. The decreasing productivity will lead to slow economic growth. It will also cause air pollution when traffic congestion reduces the vehicle speed lower than 72km/h. Vehicles spending more time on the road will increase the carbon dioxide emission to the open air [2].

This project is different from other traffic data collection techniques such as Automated Vehicle Identification systems, video cameras, inductive loops and radar-based sensors. The innovativeness of this project is it utilizes the onboard Global Positioning System (GPS) in each city bus such as Rapid Kuantan in Pahang state, Malaysia. Geolocation and time information is provided to a GPS receiver in all weather conditions and any places where line-of-sight to the GPS satellites acts as moving sensors travelling in certain routes and does not require to set up any instrumentation on the roadway. It also can provide accurate position and velocity information so it can be used to obtain the traffic information by estimating the traffic condition of the route [3, 4, 5]. Further, the amount of bus which has onboard GPS keeps increasing these years for tracking purpose. So, accurate traffic congestion can be predicted by using the onboard GPS data and it is cost saving compare to those fixed-point traffic sensors such as cameras where high installation and maintenance cost are needed.

The project implementation includes the collection of city bus GPS data and then the algorithm to determine the traffic condition according to the bus speed is developed using Fuzzy Logic Algorithm in MATLAB. At the final phase of the project, an Android Application is developed to display the traffic condition and push a notification to the user when the travelling route faced a heavy traffic condition.

II. RELATED WORKS

Most traffic condition estimation and monitoring are based on camera imaging. Projects by Posawang et al. and Kumar et al. are using ANN to determine the road traffic congestion based on the image processing information [6, 7].

Posawang et al. classified the level of traffic congestion based on motorists' perceptions by imitating their visual judgments automatically using the image processing information obtained from traffic cameras every minute [6]. The web survey is then implemented by using Google Map APIs, PHP, AJAX and the PostgresSQL database. Whereas, Kumar et al. used past traffic data which using video cameras at selected locations for a period of five days from Monday to Friday between 11.00 am to 1.00pm and applied it in the ANN for short-term prediction of traffic flow in non-urban highway where the input variables were traffic volume, speed density, time and day of the week [7]. A large number of samples were used to train the data by using Multilayer-Perceptron Neural Network Model (MLP) which is a feedforward artificial neural network model and consists of multiple layers of nodes for both projects [6, 7]. MLP has any number of inputs and has one or more hidden layers with any number of units. It uses linear combination functions in the input layer, sigmoid activation function, has connections between the input layer and the first hidden layer and between the last hidden layer and the output layer [8].

In brief, vehicle speed can be used to classify the levels of traffic congestion for the data collected from various kinds of sensors other than traffic camera since it was the factor which most affected by the congestion [6]. Besides that, the ANN model also could predict vehicle speed and count accurately and its performance also will be consistent even if the time interval for the prediction is increased [6, 7]. However, the cause and effect of each of independent variable can be found in fuzzy logic but ANN framework is impossible to find the mutual interrelation between the variables [7]. The ANN performance also mainly depends on the network training and is extremely by the size and quality of training data and different design parameters like the number of hidden layers, number of epoch and the number of elements in each hidden layer [10].

Projects done by Shankar et al., Eze et al. and Sharma et al. used Fuzzy Logic to determine the traffic condition because the fuzzy logic model is able to produce stable traffic volume predictions with limitations of fuzzy control rules on a small dataset. [9, 10, 11].

Shankar et al. were also using video imaging as fuzzy input to estimate the road traffic congestion of a road segment in Dehradun city [9]. Whereas, Eze et al. were using fuzzy logic to control the length of the green light time which depends on the traffic conditions [11]. There are two electromagnetic sensors placed on the road for each lane in this system which could drastically increase the real implementation cost. Sharma et al. on the other hand were using fuzzy logic to predict the traffic volume in weekdays at Delhi-hardwar highway (NH-58) road and the results are promising with an average of MAPE error is 15% [10].

Therefore, for this project where a small dataset is used, fuzzy logic is a potential method in representing human behaviour knowledge and in reasoning with that knowledge to make useful inferences or actions in the estimation of traffic conditions [9, 10, 11].

III. METHODOLOGY

A. *System Function Flow*

The function flow of the system is illustrated in Figure 1. Firstly, a fuzzy logic inference system (FIS) model is generated by using MATLAB Fuzzy Logic Toolbox. Next, the recorded GPS data is loaded and sent as the first input of the FIS model. At the same time, bus dwell time was also calculated whenever the bus speed was lower than 2km/h and sent as a second input of the FIS model.

Then, if the output of the FIS model is matched with the expected result, then the output will be used as the input into the MATLAB GUI. Otherwise, if the output of the FIS model does not match with the expected result then a new FIS model will be generated again to get the best result. Finally, the congested segment of the route will be displayed in the MATLAB GUI.

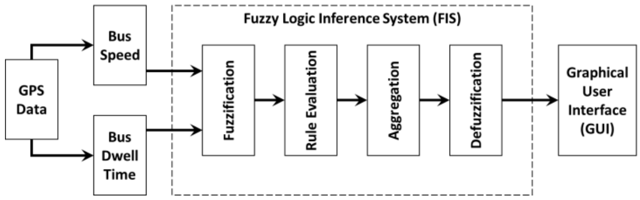

Figure 1: Block diagram of traffic congestion estimation system

B. *Data Acquisition*

The average city bus speed, maximum allowable city bus speed and the average dwell time at each bus stop are acquired from the RapidKuantan Bus Service. The GPS data is recorded by using two different devices for the verification process. The first device is the Android Mobile A-GPS (Assisted Global Positioning System) and the second device is the GPS data logger hardware developed using Arduino UNO R3, Cytron SKM53 GPS Shield and SD-card Shield V4 as shown in Figure 2. This hardware can be acting as a standalone system and the bus GPS data can be straight away logged into the SD card. The SKM53 GPS module is used because of its high performance of navigation even in harsh GPS visibility environments.

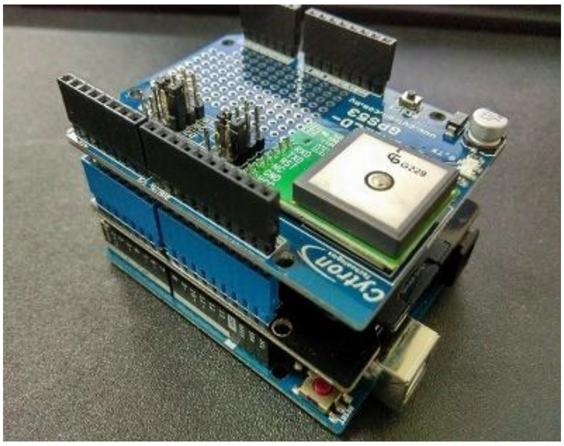

Figure 2: GPS Data Logger Hardware

The GPS data was collected on RapidKuantan Buses and Selangor Smart Buses. Six different routes are focused on this project as shown in Table 1. Six routes and two different buses company are chosen to prove the robustness of the developed system. Besides that, handouts are given to the bus users for marking the location of moderate traffic and heavy traffic based on their perspective for results verification later. A total of 145 GPS data were collected in this project to be tested in the developed algorithm.

Table 1
Routes of Buses Focused on This Project

Bus Service	Route	Total Journey
RapidKuantan	UMP – Pekan Town	32
	UMP – Kuantan	17
Selangor Smart Bus	Sungai Chua – Kajang Town	59
	Sungai Chua – Bandar Baru Bangi	33
	Sungai Chua – Seri Kembangan	2
	Sungai Chua – Petaling Jaya	2
Total		145

C. *Traffic Conditions Algorithm Development Using MATLAB R2016B*

The algorithm is developed using MATLAB to predict the traffic condition. MATLAB is used to extract the bus speed information and time from the collected GPS data. Then the bus dwell time is calculated when the bus speed is less than 2km/h to differentiate between the heavy traffic condition and bus stopping at the bus stop or traffic lights.

Before calculating the bus dwell time, the GPS data is divided into 4 groups. Group 3 is the bus speed more than 30km/h, Group 2 is the bus speed between 30km/h and 10km/h, Group 1 is the bus speed between 10km/h and 2km/h and Group 0 is the bus speed less than 2km/h. This grouping is done based on the average speed at different traffic conditions as in Table 2.

Table 2
Average Speed for Traffic Condition Estimation

Condition	Urban [12]	Highway [6]	This Project
Heavy	< 5 km/h	< 17 km/h	< 10 km/h
Moderate	5 – 10 km/h	17 – 30 km/h	10 – 30 km/h
Clear	> 10 km/h	> 30 km/h	> 30 km/h

The Group 0 is used to calculate the bus dwell time. The bus dwell time is the difference between maximum time and minimum time in the group 0. This is done by using MATLAB function of intmax which is the largest value of the integer is used as the reference to find out the minimum time correctly.

Fuzzy Logic Toolbox of the MATLAB is used to generate a Fuzzy Logic Inference System (FIS) model for the project. The FIS model can be created by using graphical tools functions of this software which is Fuzzy Logic Designer as shown in Figure 3. The Mamdani model is used to develop the FIS model that consists of 4 steps which are fuzzification, rule evaluations, rule aggregation and defuzzification. The accuracy of Mamdani model and Sugeno model are similar but the Mamdani model is used in this project because the Mamdani model may improve the overall performance [9, 13]. The bus speed and bus dwell time are created in this FIS model as the input variables and the output is the level of congestion (LOC).

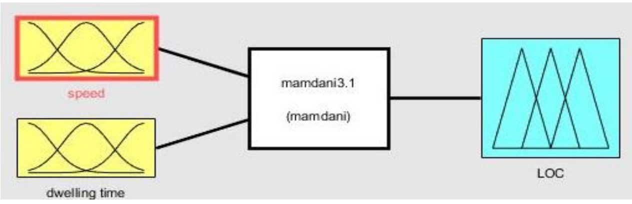

Figure 3: Fuzzy Logic Designer

The first input is the bus speed. The variables of this membership functions are set as very slow, slow, medium and fast as shown in Figure 4. The range set for very slow is between 0 km/h to 3.8 km/h and for slow is between 3.6 km/h to 14 km/h. Then, 9.5 km/h to 35 km/h is the range set for medium and more than 26.2km/h is the range of fast. This range of intersection is set because there are cases where speed is between two conditions. For example, it can be considered as partially medium or slow in certain speed as the vehicle might just speed up for a while during congestion.

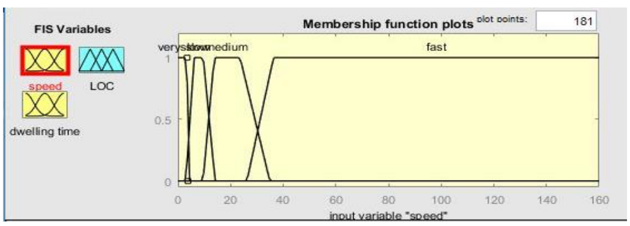

Figure 4: Parameters set for bus speed

The second input is the bus dwell time with the variables of short and long as shown in Figure 5. The range set for short is below than 5 minutes whereas the range set for long is above 3 minutes. This range is chosen because the average dwell time for buses where the fare is collected by using cash such as in the case of most bus user in Malaysia is 168 seconds or nearly 3 minutes [16]. This range of intersection is also set to consider the duration the bus stops at traffic lights. Besides that, the type of the membership functions for the inputs are all being set as Trapezoidal-shaped and intersect with each other [9, 14, 15]. This is because it will get the best results which match with condition created.

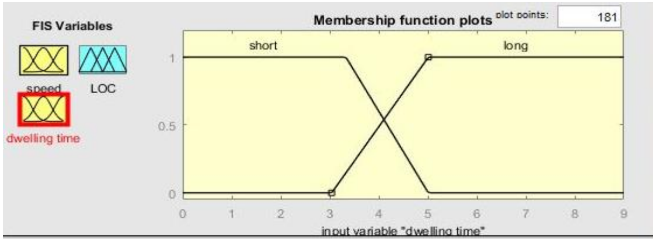

Figure 5: Parameters set for bus dwell time

The output is the level of congestion (LOC) with the variables of heavy, moderate and clear as shown in Figure 6. The membership functions of LOC is in triangular-shaped [10, 15]. The three membership functions are not interacting with each other. From the figure, the range for heavy LOC is set between 0 to 2, the range for moderate LOC is set from 2 to 4 and the range for clear LOC is from 4 to 6. These values represent different traffic conditions and needed for output calculations later. By using this FIS model, the output after inserting the manually adjusted crisp inputs are the best fit to the planned conditions. Both triangular and trapezoidal-shaped are simple to implement and fast for computation. Basically, these parameters are the points of the universe of discourse for which the membership function reaches its minimum and maximum values, zero and one [17]. This FIS model is saved for further use. 5 fuzzy rules were set for this FIS model. The rules set are shown in Figure 7.

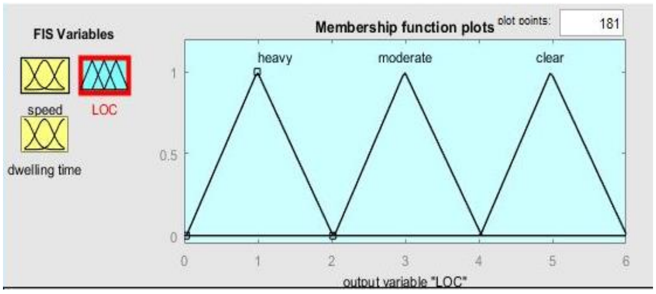

Figure 6: Parameters set for Level of Congestion (LOC)

RULE 1:
IF speed is very slow AND dwelling time is long THEN LOC is heavy

RULE 2:
IF speed is fast THEN LOC is clear

RULE 3:
IF speed is slow THEN LOC is heavy

RULE 4:
IF speed is medium THEN LOC is moderate

RULE 5:
IF speed is very slow AND dwelling time is short THEN LOC is clear

Figure 7: Fuzzy rules set for FIS model

Besides the bus speed and bus dwell time, another parameter that needs to be considered for determining the traffic condition is the length of travel at certain speed range. This is to ensure the system does not directly classify any slow speed as congestion or medium speed as clear. It must travel with slow speed at a certain distance before it could be classified as congestion and vice versa. Therefore, the minimum distance of travel at certain speed range is required. The minimum distance travel can be calculated from the Equation (1) where L is the length of travel in meter, n is the typical number of many cars queuing during red traffic light which is 12 cars [18], d_c is the stop distance between each vehicle in queues which is 2m [19] and l_c is the average length for light vehicles such as cars, taxis, vans and multi-purpose vehicle (MPV) which is 5.5m [20]. The calculated minimum distance travel is 90m.

$$L = n \times (l_c + d_c) \quad (1)$$

When the travel distance at certain speed range does not exceed the minimum travel distance then it will maintain the classification as the previous segment. For example, if the previous decision is clear traffic, it will maintain clear traffic in the next road segment. But, if the previous decision is heavy traffic, it will maintain heavy traffic in the next road segment.

IV. RESULTS AND DISCUSSION

A. *Graphical Interface*

A Graphical User Interface (GUI) is developed in MATLAB to show the graphical representation of the results obtained from the developed traffic condition monitoring algorithm. The heavily congested part is shown in red colour, the moderately congested part is shown in yellow colour and the clear traffic part is shown in green colour. Everything will be shown on google map by using Google Map Application Programming Interfaces (API) in MATLAB.

Whenever the heavily congested road segments appear in the GUI, the heavily congested road address will be shown in the text box. This is done by building the URL based on the longitude and latitude. This can build an XML document which consists of the nested node with the locations of the GPS coordinate. Then, the response and the relevant information is extracted by using the xmlread function of the MATLAB. This xmlread function will read the XML document and return Document Model node.

There is also a push notification button which can send the heavily congested data into a google spreadsheet and then push a notification to the Android phone which is developed using MIT App Inventor. This is done by using two function files called as RunOnce and mat2sheets which can be downloaded from the file exchange MATLAB central. These functions enabling the user to send data to google spreadsheets by enabling the Google Drive and Sheets APIs. Then, credentials required to create via an OAuth 2.0 client ID that comes with a Client ID and Client Secret. These are the two arguments that are required to input in the code of RunOnce function file. The RunOnce function file needs to run first before the mat2sheets function file as it will request the security token from the Google Drive.

The push notification is included in this project to alert the user about which road is heavily congested so that he or she can avoid using that road. If there is no heavy congestion, then there will be no notification will be pushed. Besides that, web viewer is also used to show the exact location of heavily congested in Google Map so that the users can know the exact location easily. Figure 8, Figure 9 and Figure 10 shows the results shown in GUI for three different routes.

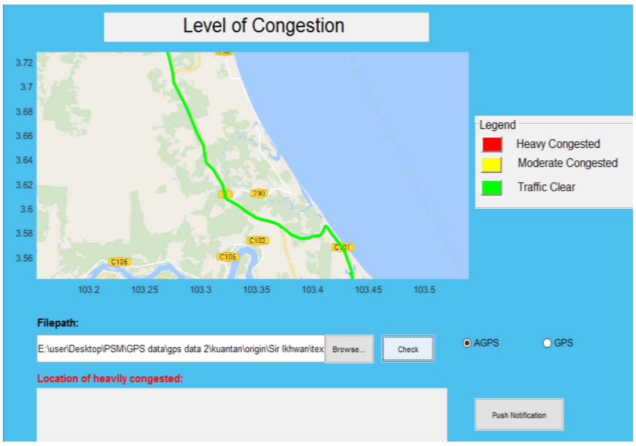

Figure 8: Results for UMP – Kuantan Route

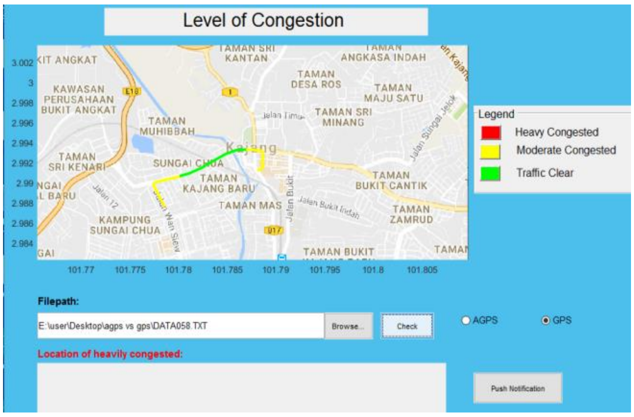

Figure 9: Results for Sungai Chua – Kajang Town Route

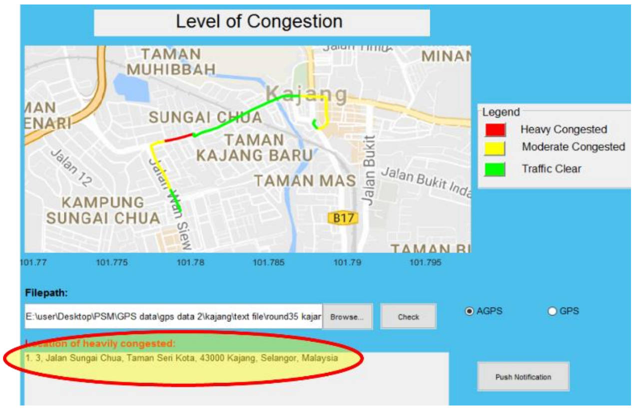

Figure 10: Results for Sungai Chua – Kajang Town Route with heavily congested area

From the results, only Figure 10 has a road segment with the heavily congested area. The address for the heavily congested area is then shown in the highlighted text box. This address does not appear in Figure 8 and Figure 9 because there is no heavily congested area appear in the results. The location of heavily congested is then sent to Google spreadsheet when the push notification button is clicked. The result is shown in Figure 11. At the same time, a notification is pushed to an android mobile to alert the user. The result is shown in Figure 12.

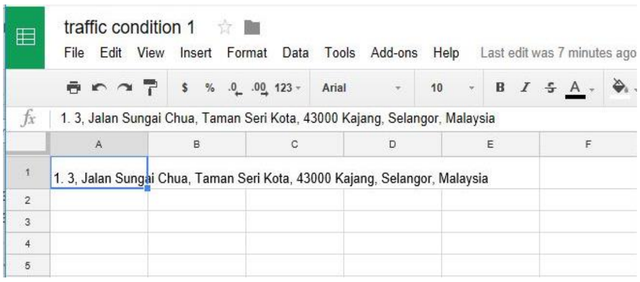

Figure 11: Google spreadsheet as database

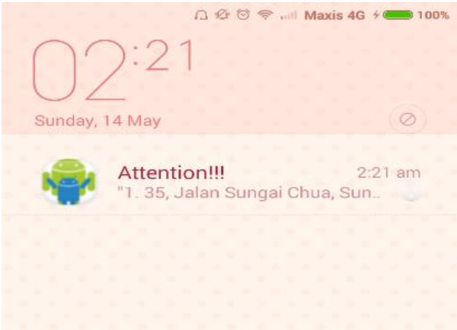

Figure 12: Push Notification received on an Android Phone

After the user clicked on the notification, then it will lead to the main page of the app as shown in Figure 13. The full address of the heavily congested location is then shown in this main page. The location of heavily congested is also shown in the Google Map so that the user can identify the congested area easily.

Figure 13: Main page of the app

B. *Results Verification*

There is a total of 145 GPS data were collected in RapidKuantan Bus and Selangor Smart Bus to verify and test the accuracy of the system. The given handouts are marked by the bus users to show the location of moderate traffic and heavy traffic based on their perspective. Other than that, a few videos are also recorded as a proof to verify the accuracy of the system.

After comparing the results in the GUI with the handouts, most of the data is accurate but there is also inaccuracy appeared in a few data. The results after verification are shown in Table 3.

Table 3
Data Verification

Bus Service	Route	Total Journey	Accurate Results
RapidKuantan	UMP – Pekan Town	32	27
	UMP – Kuantan	17	12
Selangor Smart Bus	Sungai Chua – Kajang Town	59	59
	Sungai Chua – Bandar Baru Bangi	33	29
	Sungai Chua – Seri Kembangan	2	2
	Sungai Chua – Petaling Jaya	2	2
Total		145	131

From the Table 3, the accuracy of this system is 90.34%. The inaccurate data occurred mostly in the data from UMP to Kuantan and UMP to Pekan Town. This is due to the unexpected road conditions exist in this route during the data collection such as the deflection of the road, sharp corner, uneven road surface, holes and wild animals crossing the road which cause the bus driver to slow down the speed and even stop for a while.

V. CONCLUSION

Road traffic condition is successfully predicted by using Fuzzy logic in this project. Three parameters are considered as the fuzzy input to the traffic condition estimation algorithm which are the bus speed, bus dwell time and minimum travel distance at a certain speed range. Due to the limitations of datasets available, fuzzy logic is the most suitable algorithm to be used in this project. The project is considered a success as the accuracy of this system is 90.34% and the heavy traffic location can be pushed automatically to the user.

For further improvement, other parameters can be considered and add to the algorithm such as random speed changing behaviour to reduce the effect of the road deflection, sharp corner, uneven road surface, holes and wild animals crossing the road. Recording GPS data of multiple city buses that travel at the same route is also important to improve the accuracy of the system.

ACKNOWLEDGMENT

This research is supported by Universiti Malaysia Pahang Internal Grant of RDU1703164. The authors would also like to thank the Faculty of Electrical & Electronics Engineering Universiti Malaysia Pahang for providing the facilities to conduct this research and financial support throughout the process.

REFERENCES

[1] P. Asier, H.-J. Unai, O. Enrique and G.-Z. Ignacio Julio, *Intelligent Transport Systems: Technologies and Applications*, United Kingdom: John Wiley & Sons Ltd, 2016.

[2] Matthew Barth and Kanok Boriboonsomsin, “Traffic Congestion and Greenhouse Gases,” in *Access*, vol. 35, 2009, pp. 2-9.

[3] Elliott D. Kaplan and Christopher J. Hegarty, *Understanding GPS/GNSS: Principles and Applications*, 3rd ed., London: Artech House, 2017.

[4] Arijit Chowdhury, Tapas Chakravarty and P. Balamuralidhar, “Estimating True Speed of Moving Vehicle using Smartphone-based GPS Measurement,” in *IEEE International Conference on Systems, Man, and Cybernetics*, 2014, pp. 3348-3353.

[5] J. Mohamad and A. T. Kinggundu, “The Rise of The Private Car in Kuala Lumpur, Malaysia – Assessing the Policy Options –,” in

International Association of Traffic and Safety Sciences Research, Vol. 31, No. 1, 2007, pp. 69-77.

[6] Pitiphoom Posawang, Satidchoke Phosaard, Weerapong Polnigongit and Wasan Pattara-Atikom, "Perception-based Road Traffic Congestion Classification using Neural Networks and Decision Tree," in *Electronic Engineering and Computing Technology*, Springer Netherlands, 2010, pp. 237-248.

[7] Kranti Kumar, M. Parida and V.K. Katiyar, "Short Term Traffic Flow Prediction for a Non-Urban Highway Using Artificial Neural Network," in *2nd Conference of Transportation Research Group of India (2nd CTRG)*, 2013, pp. 755-764.

[8] G.K. Jha, *Artificial neural networks and its applications*, New Delhi: IARI, 2007.

[9] Hari Shankar, P. L. N. Raju and K. Ram Mohan Rao, "Multi Model Criteria for the Estimation of Road Traffic Congestion from Traffic Flow Information Based on Fuzzy Logic," *Journal of Transportation Technologies*, vol. 2, 2012, pp. 50-62.

[10] Bharti Sharma, Vinod Kumar Katiyar and Arvind Kumar Gupta, "Fuzzy Logic Model for the Prediction of Traffic Volume in Week Days," *International Journal of Computer Applications*, vol. 107, no. 17, 2014, pp. 1-6.

[11] U.F. Eze, Igoh Emmanuel and Etim Stephen, "Fuzzy Logic Model for Traffic Congestion," *IOSR Journal of Mobile Computing & Application*, vol. 1, no. 1, 2014, pp. 15-20.

[12] W. Pattara-atikom, P. Pongpaibool, and S. Thajchayapong, "Estimating Road Traffic Congestion using Vehicle Velocity," in *6th International Conference on ITS Telecommunications Proceedings*, 2006, pp. 1001-1004.

[13] Yang Wang and Yanyan Chen, "A Comparison of Mamdani and Sugeno Fuzzy Inference Systems for Traffic Flow Prediction," *Journals of Computers*, vol. 9, no. 1, 2014, pp. 12-21.

[14] Jia Lu and Li Cao, "A Quantization Method of Traffic Congestion Evaluation Based on Fuzzy Logic," in *IEEE International Conference on Systems, Man and Cybernetics*, 2003, pp. 3348-3351.

[15] Panita Pongpaibool, Poj Tangamchit and Kanokchai Noodwong, "Evaluation of Road Traffic Congestion Using Fuzzy Techniques," in *TENCON 2007 - 2007 IEEE Region 10 Conference*, 2007, pp. 1-4.

[16] M.H. Hafezi, A. Ismail and A.A. Shariff, "Comparative Analysis of Fare Collection System on Bus Operations," *Journal of Applied Sciences*, vol. 12, no. 4, 2012, pp. 393-397.

[17] B. Ayyub. Ed, *Applied research in uncertainty modeling and analysis*, Vol. 20, Springer Science & Business Media, 2007.

[18] Md Anwarul Azim and Md. Nazmul Huda, "Fuzzy Traffic Control System," in *Term Paper Based on Case Study and Implementation of a Fuzzy Application*, Queen's University.

[19] Wong Poh Lee, Mohd. Azam Osman, Abdullah Zawawi Talib, and Ahmad Izani Md. Ismail, "Dynamic Traffic Simulation for Traffic Congestion Problem Using an Enhanced Algorithm," *World Academy of Science, Engineering and Technology*, vol. 45, 2008, pp. 271-278.

[20] Ministry of Works and Transport, *Traffic Data Collection and Analysis*, Gaborone, Botswana, 2004.